ISTITUTO ITALIANO DI TECNOLOGIA PATTERN ANALYSIS AND COMPUTER VISION

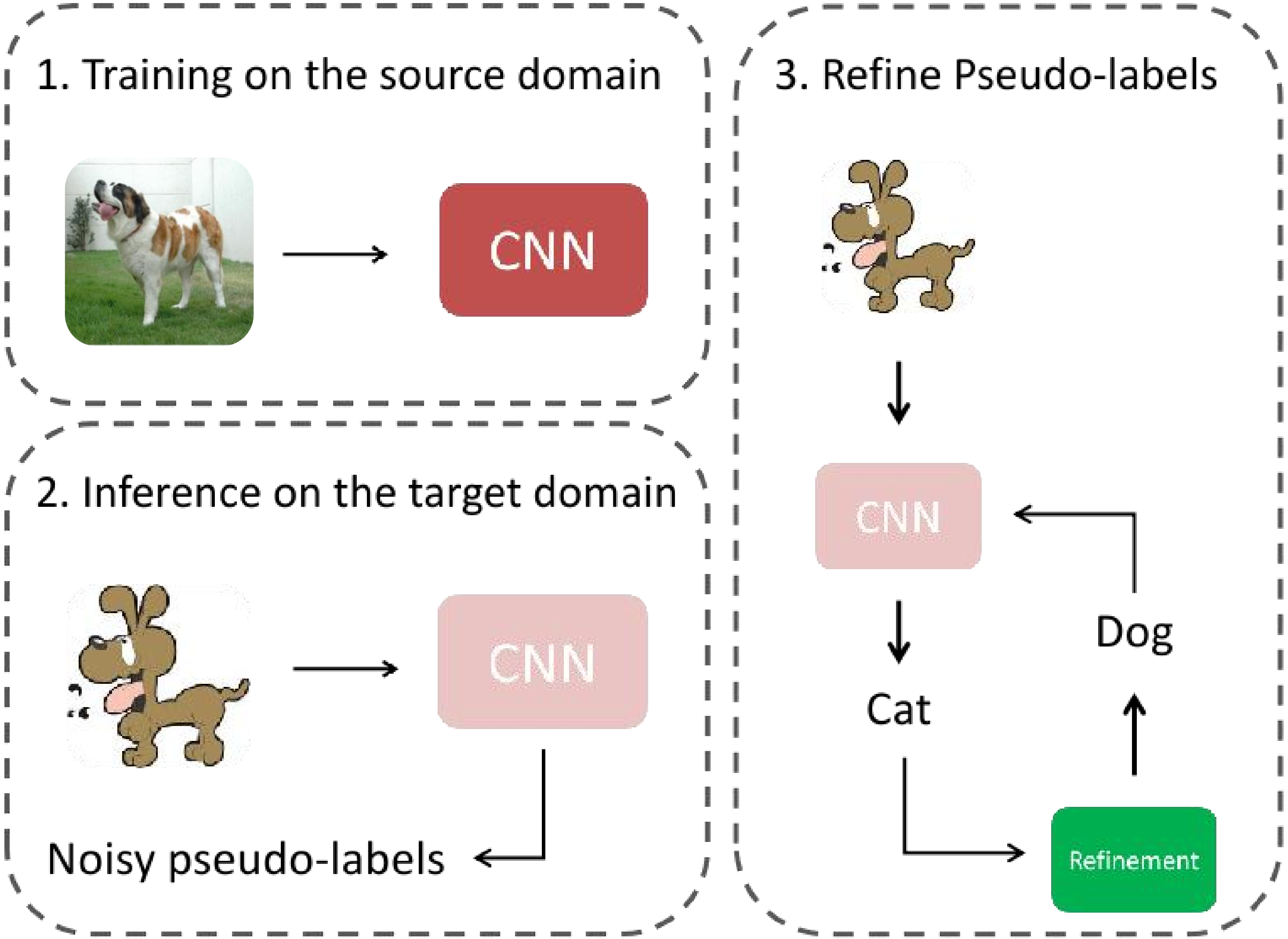
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Guiding Pseudo-labels with Uncertainty Estimation for Source-free Unsupervised Domain Adaptation

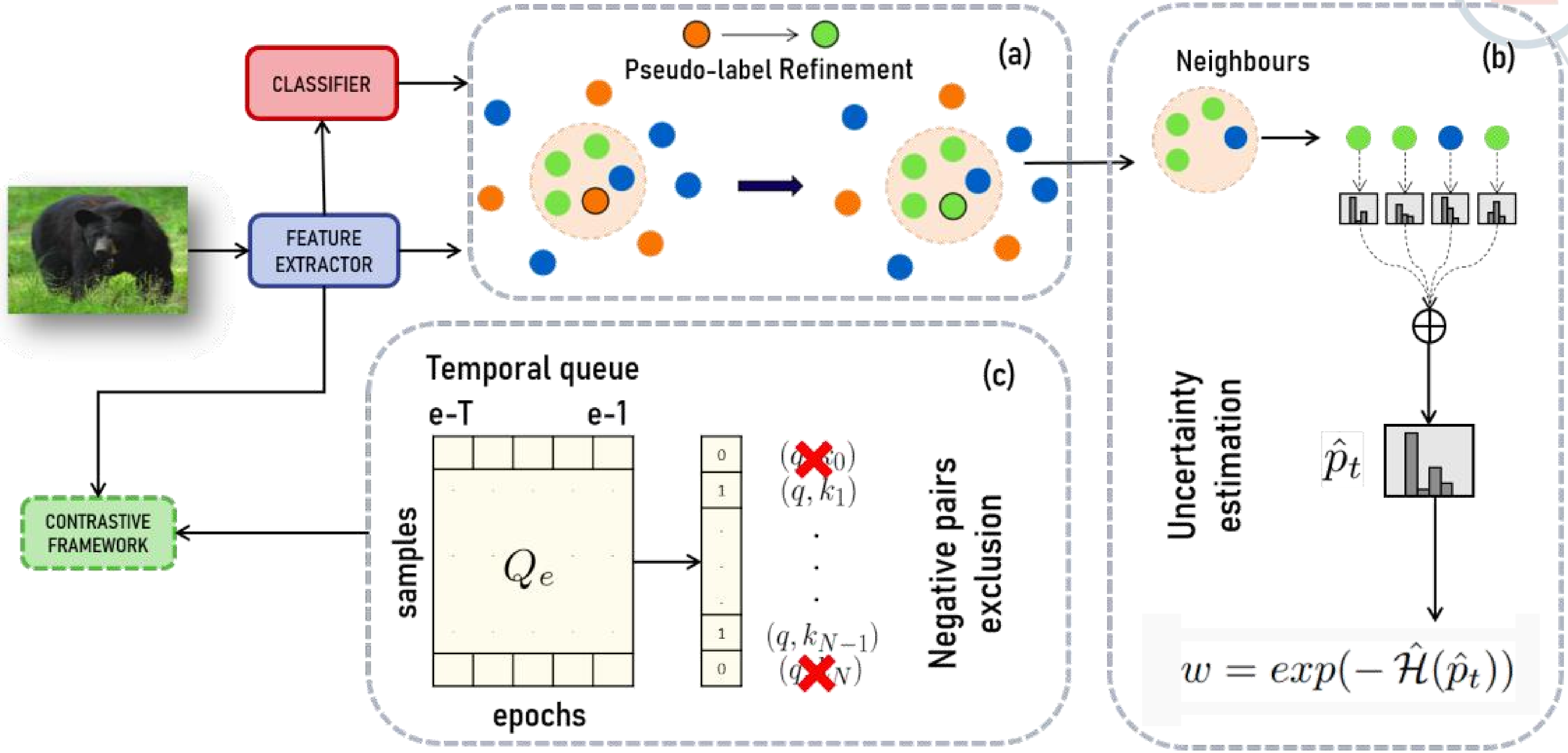
Mattia Litrico Alessio Del Bue Pietro Morerio
Italian Institute of Technology (IIT)

TUE-PM-336

Source-free Unsupervised Domain Adaptation

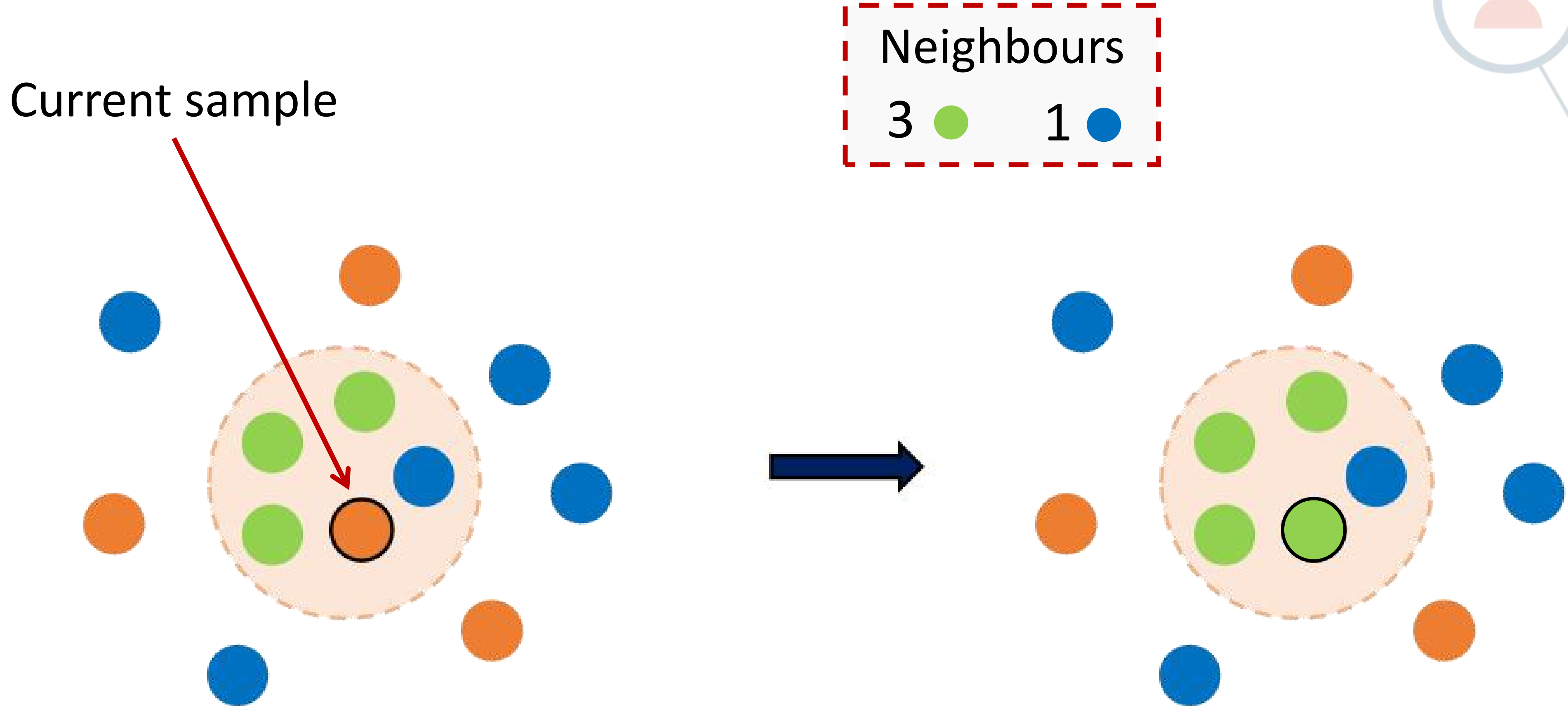


Overview



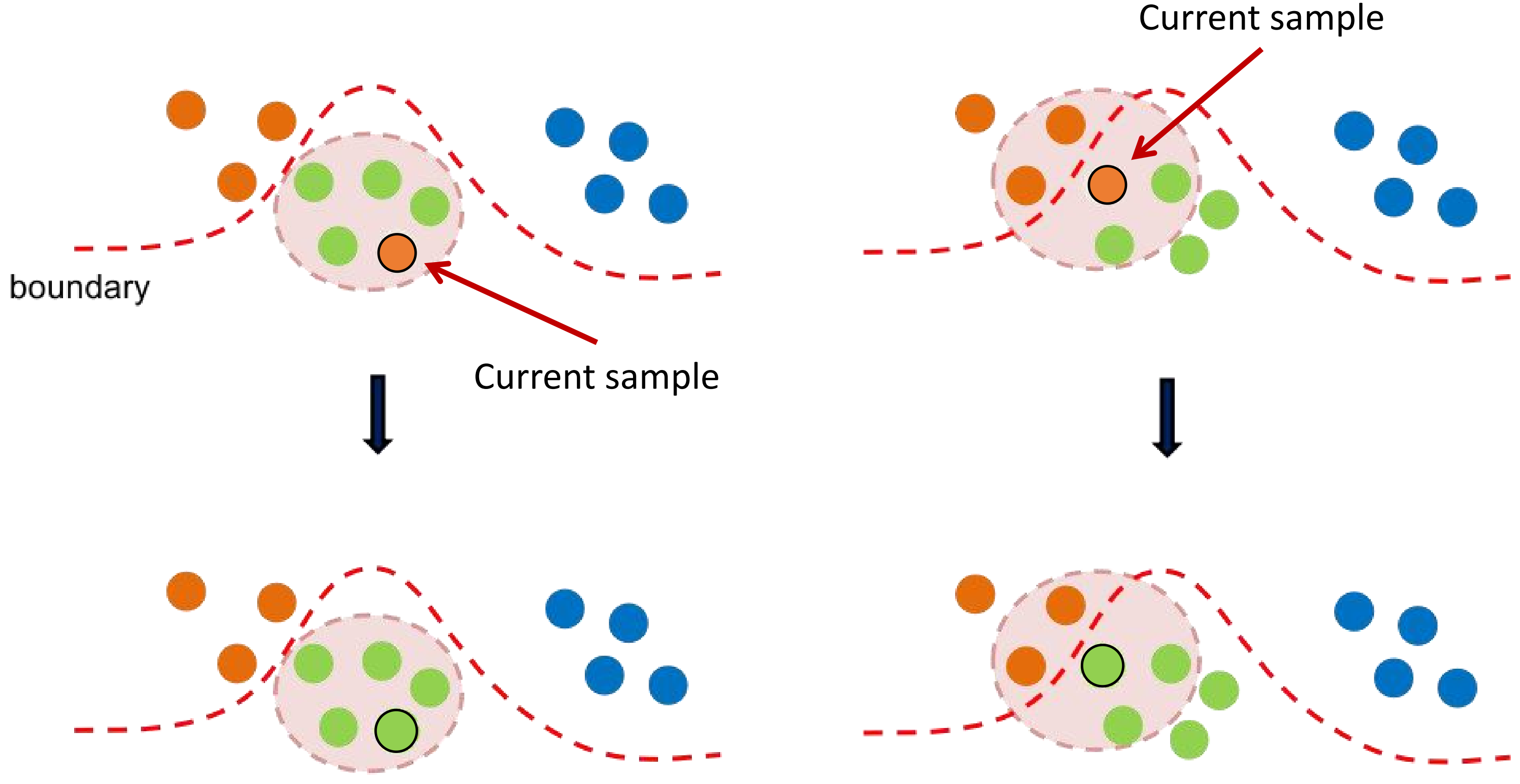
CLASSIFIER
FEATURE EXTRACTOR
CONTRASTIVE FRAMEWORK
Pseudo-label Refinement
(a)
Neighbours
(b)
Uncertainty estimation
$\hat{p}_t$
$w = exp(-\hat{\mathcal{H}}(\hat{p}_t))$
Temporal queue
(c)
e-T
e-1
samples
Q_e
epochs
0
1
1
0
(q, k_1)
(q, k_{N-1})
Negative pairs exclusion

Pseudo-label Refinement via Nearest-neighbours Knowledge Aggregation



Pseudo-labels are refined by aggregating knowledge from neighbours samples using a voting strategy on their predictions.

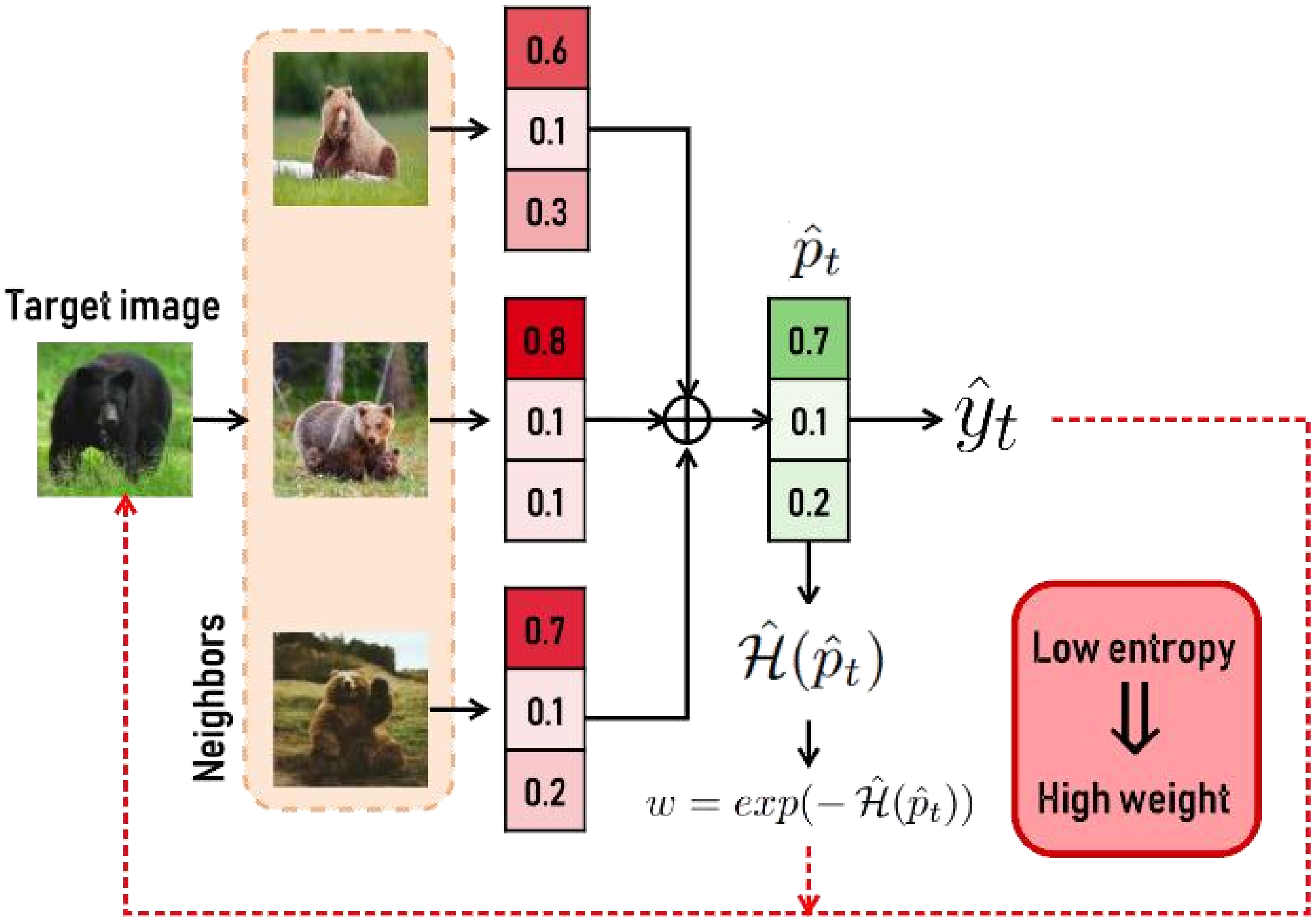
Pseudo-label Refinement via Nearest-neighbours Knowledge Aggregation



Low Uncertainty

High Uncertainty

Loss Reweighting with Pseudo-labels Uncertainty Estimation



The entropy of the averaged neighbours' predictions measures the uncertainty of the pseudo-label and the corresponding weight in the classification loss.

Standard contrastive learning framework ([1], [2])

Negative Pair

Current pseudo-label

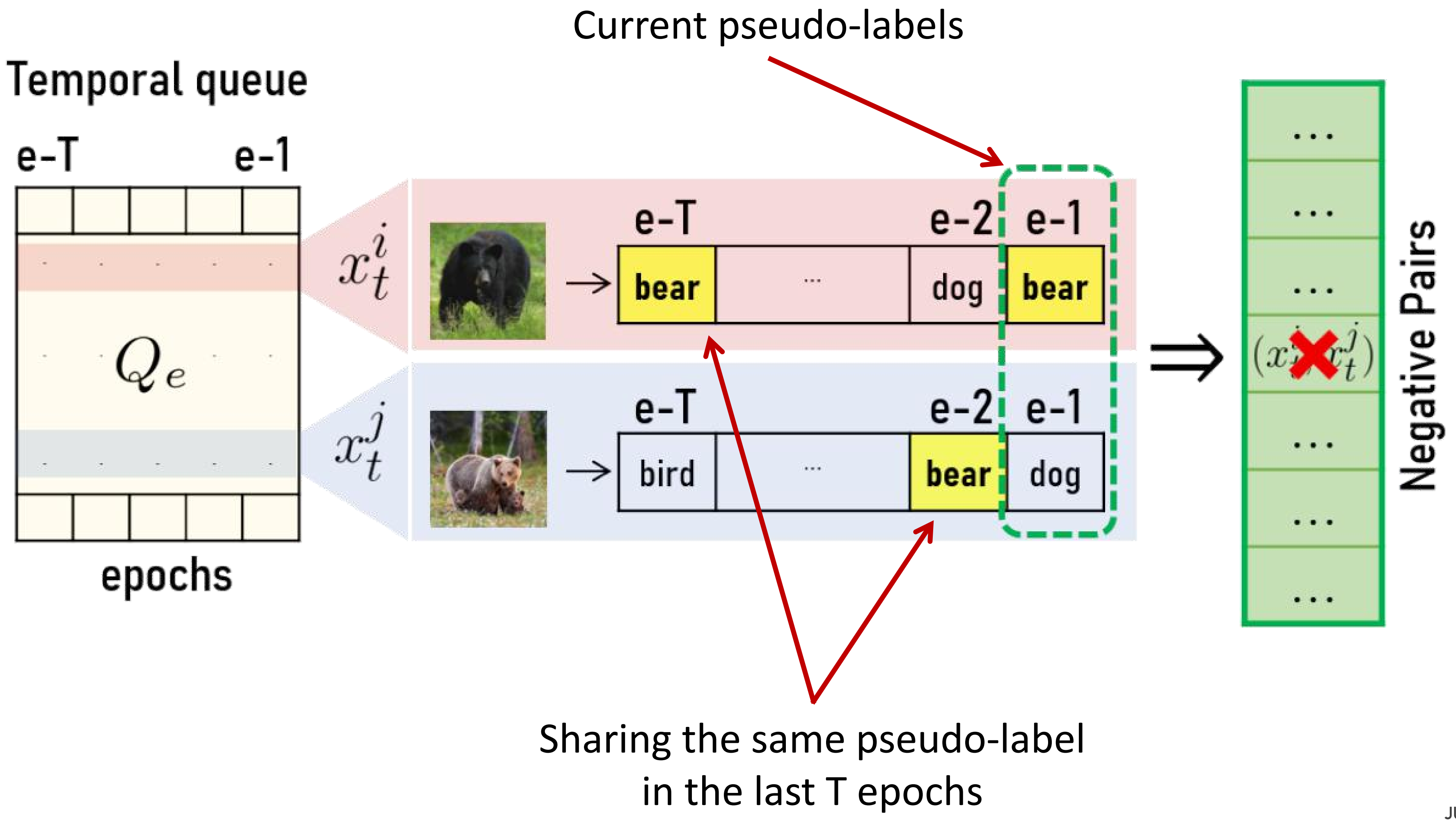
 Bear

 Dog

Negative Pair

[1] Chen et al., “A Simple Framework for Contrastive Learning of Visual Representations”, ICML 2020
[2] Chen et al., “Contrastive Test-Time Adaptation”, CVPR 2022

Temporal Queue for Negative Pairs Exclusion



Joint training with self-learning

$$L_t = \gamma_1 \boxed{L_t^{cls}} + \gamma_2 \boxed{L_t^{ctr}} + \gamma_3 \boxed{L_t^{div}}$$

Negative learning classification loss.

$$L_t^{cls} = -\,\mathbb{E}_{x_t \in \mathcal{X}_t} \left[w_{x_t} \cdot \sum_{c=1}^{C} \tilde{y}^c \, log \, (1 - p_{sa}^c) \right]$$

Contrastive loss.

$$L_t^{ctr} = L_{\text{InfoNCE}} = -\,log \, \frac{exp(q \cdot k_+ / \tau)}{\sum_{j \in \mathcal{N}_q} exp(q \cdot k_j / \tau)}$$

$$\mathcal{N}_q = \{ j | \hat{y}_j^i \neq \hat{y}^i, \; \forall j \in \{1, ..., N\}, \forall i \in \{1, ..., T\} \}$$

Divergence loss.

$$L_t^{div} = \mathbb{E}_{x_t \in \mathcal{X}_t} \sum_{c=1}^{C} \bar{p}_q^c \, log \, \bar{p}_q^c, \quad \bar{p}_q = \mathbb{E}_{x_t \in \mathcal{X}_t} \, \sigma(g_t(t_{sa}(x_t)))$$

Results

Method	SF-UDA	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Avg.
CDAN [42]	no	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
CDAN+BSP [9]	no	92.4	61.0	81.0	57.5	89.0	80.6	90.1	77.0	84.2	77.9	82.1	38.4	75.9
SWD [30]	no	90.8	82.5	81.7	70.5	91.7	69.5	86.3	77.5	87.4	63.6	85.6	29.2	76.4
MCC [25]	no	88.7	80.3	80.5	71.5	90.1	93.2	85.0	71.6	89.4	73.8	85.0	36.9	78.8
CAN [26]	no	**97.0**	87.2	82.5	74.3	**97.8**	96.2	90.8	80.7	96.6	**96.3**	87.5	59.9	87.2
DivideMix [35]	yes	95.0	82.4	85.3	78.1	94.2	90.3	90.1	81.3	92.5	91.9	91.2	60.8	86.1
MA [36]	yes	94.8	73.4	68.8	74.8	93.1	95.4	88.6	84.7	89.1	84.7	83.5	48.1	81.6
BAIT [37]	yes	93.7	83.2	84.5	65.0	92.9	95.4	88.1	80.8	90.0	89.0	84.0	45.3	82.7
SHOT [38]	yes	95.3	87.5	78.7	55.6	94.1	94.2	81.4	80.0	91.8	90.7	86.5	59.8	83.0
DIPE [64]	yes	95.2	87.6	78.8	55.9	93.9	95.0	84.1	81.7	92.1	88.9	85.4	58.0	83.1
NEL [1]	yes	94.5	60.8	**92.3**	87.3	87.3	93.2	87.6	**91.1**	56.9	83.4	93.7	**86.6**	84.2
A^2 Net [67]	yes	94.0	87.8	85.6	66.8	93.7	95.1	85.8	81.2	91.6	88.2	86.5	56.0	84.3
G-SFDA [68]	yes	96.1	88.3	85.5	74.1	97.1	95.4	89.5	79.4	95.4	92.9	89.1	42.6	85.4
SFDA-DE [11]	yes	95.3	**91.2**	77.5	72.1	95.7	**97.8**	85.5	86.1	95.5	93.0	86.3	61.6	86.5
AdaContrast [5]	yes	**97.0**	84.7	84.0	77.3	96.7	93.8	**91.9**	84.8	94.3	93.1	**94.1**	49.7	86.8
CoWA [32]	yes	96.8	90.3	87.0	67.4	97.2	96.6	90.4	87.3	95.6	95.5	91.8	62.5	88.2
Ours	yes	97.3	96.2	90.5	**91.8**	90.0	94.2	87.4	87.7	**97.0**	84.3	93.0	81.0	**90.0**

Ablation studies

Pseudo-label refinement	Contrastive regularisation	Negative learning	Temporal-queue exclusion	Uncertainty reweighting	Avg. Acc.
✓	✗	✗	✗	✗	52.3
✓	✓	✗	✗	✗	78.9
✓	✓	✓	✗	✗	82.1
✓	✓	✓	✓	✗	85.8
✓	✓	✓	✓	✓	90.0

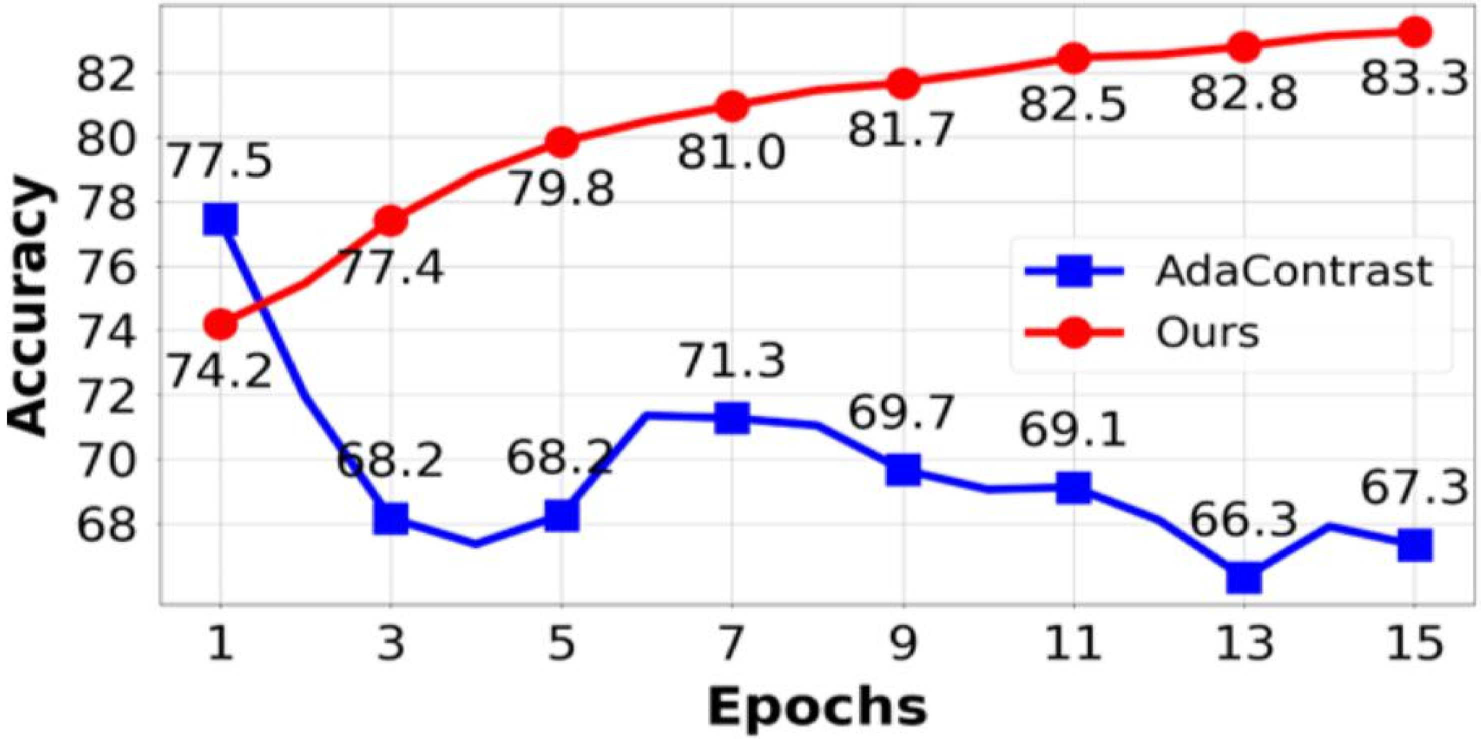
Ablation studies of sub-components.

Method	Acc.
Ours w/ hard entropy margin	85.9
Ours w/ linear weighting	85.1
Ours w/ positive	83.0
Ours w/ positive+negative	85.2
Ours	**90.0**

Additional analysis.

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Refining Pseudo-labels during the adaptation



Our method guides the pseudo-labels refinement and mitigates the effects of noisy samples, resulting in progressively improving the pseudo-labels accuracy.

Chen et al., "Contrastive Test-Time Adaptation", CVPR 2022

Project page and code

https://github.com/MattiaLitrico/Guiding-Pseudo-labels-with-Uncertainty-Estimation-for-Source-free-Unsupervised-Domain-Adaptation